BEYOND ℓ_1-NORM MINIMIZATION FOR SPARSE SIGNAL RECOVERY

Hassan Mansour
University of British Columbia, Vancouver - BC, Canada

ABSTRACT

Sparse signal recovery has been dominated by the basis pursuit denoise (BPDN) problem formulation for over a decade. In this paper, we propose an algorithm that outperforms BPDN in finding sparse solutions to underdetermined linear systems of equations at no additional computational cost. Our algorithm, called WSPGL1, is a modification of the spectral projected gradient for ℓ_1 minimization (SPGL1) algorithm in which the sequence of LASSO subproblems are replaced by a sequence of weighted LASSO subproblems with constant weights applied to a support estimate. The support estimate is derived from the data and is updated at every iteration. The algorithm also modifies the Pareto curve at every iteration to reflect the new weighted ℓ_1 minimization problem that is being solved. We demonstrate through extensive simulations that the sparse recovery performance of our algorithm is superior to that of ℓ_1 minimization and approaches the recovery performance of iterative re-weighted ℓ_1 (IRWL1) minimization of Candès, Wakin, and Boyd, although it does not match it in general. Moreover, our algorithm has the computational cost of a single BPDN problem.

Index Terms— Sparse recovery, compressed sensing, iterative algorithms, weighted ℓ_1 minimization, partial support recovery

1. INTRODUCTION

The problem of recovering a sparse signal from an underdetermined system of linear equations is prevalent in many engineering applications. In fact, this problem has given rise to the field of compressed sensing which presents a new paradigm for acquiring signals that admit sparse or nearly sparse representations using fewer linear measurements than their ambient dimension [1, 2].

Consider an arbitrary signal $x \in \mathbb{R}^N$ and let $y \in \mathbb{R}^n$ be a set of measurements given by $y = Ax + e$, where A is a known $n \times N$ measurement matrix, and e denotes additive noise that satisfies $\|e\|_2 \leq \epsilon$ for some known $\epsilon \geq 0$. Compressed sensing theory states that it is possible to recover x from y (given A) even when $n \ll N$, that is, using very few measurements. When x is strictly sparse—i.e., when there are only $k < n$ nonzero entries in x—and when $e = 0$, one may recover an estimate $\hat{x}$ of the signal x by solving the constrained ℓ_0 minimization problem

$$\underset{u \in \mathbb{R}^N}{\text{minimize}} \ \|u\|_0 \text{ subject to } \ Au = y. \tag{1}$$

However, ℓ_0 minimization is a combinatorial problem and quickly becomes intractable as the dimensions increase. Instead, the convex relaxation given by the ℓ_1 minimization problem

$$\underset{u \in \mathbb{R}^N}{\text{minimize}} \ \|u\|_1 \text{ subject to } \|Au - y\|_2 \leq \epsilon \tag{BPDN}$$

also known as *basis pursuit denoise* (BPDN) [3], can be used to recover an estimate $\hat{x}$. Candés, Romberg and Tao [2] and Donoho [1] show that it is possible to recover a stable and robust approximation of x by solving (BPDN) instead of (1) at the cost of increasing the number of measurements taken.

Several works in the literature have proposed alternate algorithms that attempt to bridge the gap between ℓ_0 and ℓ_1 minimization. These include using ℓ_p minimization with $0 < p < 1$ which has been shown to be stable and robust under weaker conditions than those of ℓ_1 minimization, see [4, 5, 6]. *Weighted ℓ_1 minimization* is another alternative if there is prior information regarding the support of the signal to-be-recovered as it incorporates such information into the recovery by weighted basis pursuit denoise (w-BPDN)

$$\underset{u}{\text{minimize}} \ \|u\|_{1,\mathrm{w}} \text{ subject to } \|Au - y\|_2 \leq \epsilon, \tag{w-BPDN}$$

where $\mathrm{w} \in (0, 1]^N$ and $\|u\|_{1,\mathrm{w}} := \sum_i \mathrm{w}_i |u_i|$ is the weighted ℓ_1 norm (see [7, 8, 9]).

When no prior information is available, the *iterative reweighted ℓ_1 minimization* (IRWL1) algorithm, proposed by Candès, Wakin, and Boyd [10] and studied by Needell [11], solves a sequence of weighted ℓ_1 minimization problems with the weights $\mathrm{w}_i^{(t)} \approx 1/\left|x_i^{(t-1)}\right|$, where $x_i^{(t-1)}$ is the solution of the $(t-1)$th iteration and $\mathrm{w}_i^{(0)} = 1$ for all $i \in \{1 \ldots N\}$. More recently, Mansour and Yilmaz [12] proposed a *support driven iterative reweighted ℓ_1 minimization* (SDRL1) algorithm that also solves a sequence of weighted ℓ_1 minimization problems with constant weights $\mathrm{w}_i^{(t)} = \omega \in [0, 1]$ when i belongs to support estimates $\Lambda^{(t)}$ that are updated in every iteration. The performance of SDRL1 is shown to match that of IRWL1.

hassanm@cs.ubc.ca
The author was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development Grant DNOISE II (375142-08).

Motivated by the performance of constant weighting in the SDRL1 algorithm, we present in this paper an iterative algorithm called WSPGL1 that converges to the solution of a weighted ℓ_1 problem (wBPDN) with a two set weight vector $\mathrm{w}_\Lambda = \omega$ and $\mathrm{w}_{\Lambda^c} = 1$, where $\omega \in [0, 1]$ and Λ is a support estimate. The set Λ to which the algorithm converges is not known a priori but is derived and updated at every iteration. Our algorithm is a modification of the *spectral projected gradient for ℓ_1 minimization* (SPGL1) algorithm [13] which solves a sequence of LASSO [14] subproblems to arrive at the solution of the BPDN problem. We give an overview of the SPGL1 algorithm in section 2. In contrast, our algorithm solves a sequence of weighted LASSO subproblems that converge to the solution of the wBPDN problem with weights ω applied to a support estimate Λ. We discuss the details of this algorithm in section 3 and present preliminary recovery results in section 4 demonstrating its superior performance in recovering sparse signals from incomplete measurements compared with ℓ_1 minimization. We limit the scope of this paper to discussing the algorithm and presenting sparse recovery results and leave the analysis of the algorithm for future work.

Notation: For a vector $x \in \mathbb{R}^N$, an index set $\Lambda \subset \{1 \dots N\}$ and its complement Λ^c, let x_k and $x|_k$ refer to the largest k entries of x, $x(k)$ is the kth largest entry of x, x_Λ refers to the entries of x indexed by Λ, and $x^{(t)}$ is the vector x at iteration t.

2. THE SPGL1 ALGORITHM

In this section, we give an overview of the SPGL1 algorithm, developed by van den Berg and Friedlander [13], that finds the solution to the BPDN problem.

2.1. General overview

The SPGL1 algorithm finds the solution of the BPDN problem by efficiently solving a sequence of LASSO subproblems

$$\underset{u \in \mathbb{R}^N}{\text{minimize}} \ \|Au - y\|_2 \text{ subject to } \|u\|_1 \leq \tau \qquad (\mathrm{LS}_\tau)$$

using a spectral projected-gradient algorithm. The single parameter τ determines a Pareto curve $\phi(\tau) = \|r^\tau\|_2$, where $r^\tau = y - Ax^\tau$ and x^τ is the solution of (LS_τ). The Pareto curve traces the optimal trade-off between the least-squares fit and the one-norm of the solution.

The SPGL1 algorithm is initialized at a point $x^{(0)}$ which gives an initial $\tau_0 = \|x^{(0)}\|_1$. The parameter τ is then updated according to the following rule

$$\tau_{t+1} = \tau_t + \frac{\|r^{\tau_t}\|_2 - \epsilon}{\|A^H r^{\tau_t}\|_\infty / \|r^{\tau_t}\|_2}, \qquad (2)$$

where superscript ^{H} indicates Hermitian transpose, and $\epsilon = \|e\|_2 = \|y - Ax\|_2$. Consequently, the next iterate $x^{(t+1)}$ is given by the solution of ($\mathrm{LS}_{\tau_{t+1}}$) and the algorithm proceeds until convergence.

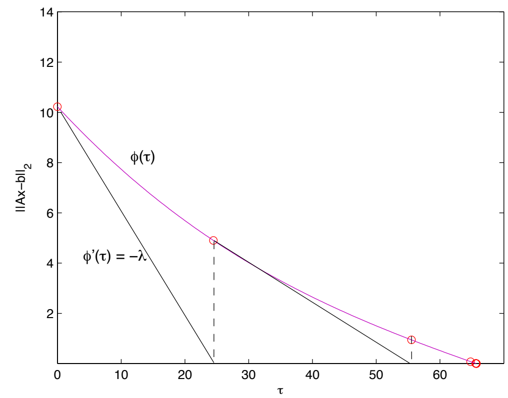


Fig. 1: Example of a typical Pareto curve showing the root finding iterations used in SPGL1 [13].

2.2. Probing the Pareto curve

One of the main contributions of [13] lies in recognizing and proving that the Pareto curve is convex and continuously differentiable over all solutions of (LS_τ). This gives rise to the update rule for τ shown in (2) and guarantees the convergence of SPGL1 to the solution of BPDN.

The update rule (2) is in fact a Newton-based root-finding method that solves $\phi(\tau) = \epsilon$. The update rule generates a sequence of parameters τ_t according to the Newton iteration

$$\tau_{t+1} = \tau_t + \frac{\epsilon - \phi(\tau_t)}{\phi'(\tau_t)}$$

where $\phi'(\tau_t)$ is the derivative of ϕ at τ_t. It is then shown that the $\phi'(\tau)$ is equal to the negative of the dual variable λ of (LS_τ) resulting in the expression $\phi'(\tau) = -\lambda = -\frac{\|A^H r\|_\infty}{\|r\|_2}$. Figure 1 illustrates an example of a Pareto curve and the root finding method used in SPGL1.

3. THE PROPOSED WSPGL1 ALGORITHM

In this section, we describe the proposed WSPGL1 algorithm for sparse signal recovery as a variation of the SPGL1 algorithm. The WSPGL1 algorithm solves a sequence of weighted LASSO subproblems to arrive at the solution to a weighted BPDN problem with weights $\omega \in [0, 1]$ applied to a support set Λ. The set Λ is derived and updated from the solutions of the weighted LASSO subproblems ($\mathrm{LS}_\tau, \mathrm{w}$).

3.1. Algorithm description

The two algorithms SPGL1 and WSPGL1 follow exactly the same initial steps until the solution x^{τ_1} of the first LASSO

problem (LS_{τ_1}) is found. At this point, WSPGL1 generates a support set Λ containing the support of the k largest in magnitude entries of x^{τ_1}. A weight vector w is then generated such that

$$\mathrm{w}_i = \begin{cases} \omega, & i \in \Lambda \\ 1, & i \in \Lambda^c \end{cases}$$

We heuristically choose $k = n/\left(2\log(N/n)\right)$ and $\omega = 0.3$.

The weight vector is then used to define the weighted LASSO subproblem

$$\underset{u \in \mathbb{R}^N}{\text{minimize}} \;\; \|Au - y\|_2 \text{ subject to } \|u\|_{1,\mathrm{w}} \leq \tau \qquad (\mathrm{LS}_{\tau,\mathrm{w}})$$

with the corresponding dual variable

$$\lambda_{\mathrm{w}} = \frac{\|A^H r\|_{\infty,\mathrm{w}}}{\|r\|_2},$$

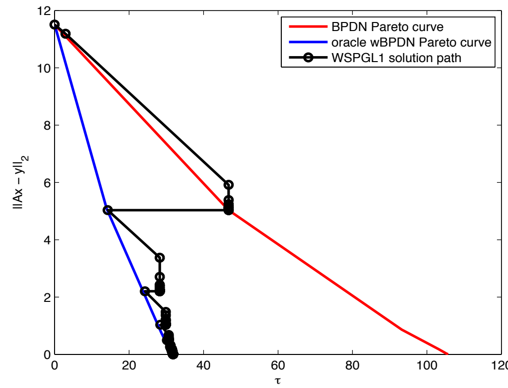
where $\|v\|_{\infty,\mathrm{w}} = \|v \cdot \mathrm{w}^{-1}\|_\infty$. The weighted LASSO subproblem and its dual constitute a subproblem of (wBPDN) with support estimate Λ. The BPDN and wBPDN problems have different Pareto curves. Therefore, the iterate $(\|r^1\|_2, \tau_1)$ which lies on the Pareto curve of BPDN must be adjusted to lie on the Pareto curve of the wBPDN problem. This can be easily achieved by switching τ_1 with $\tau_1' = \|x^{\tau_1}\|_{1,\mathrm{w}}$. The WSPGL1 algorithm then proceeds according to the following pseudocode.

Algorithm 1 The WSPGL1 algorithm

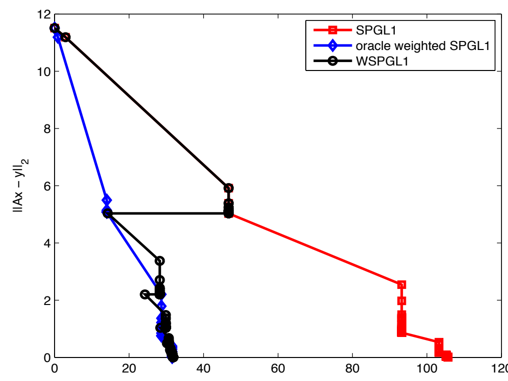
Input $y = Ax + e$, ϵ, $k = n/\left(2\log(N/n)\right)$, $\omega \in [0,1]$
Output $x^{(t)}$
Initialize $\mathrm{w}_i^{(0)} = 1$ for all $i \in \{1 \ldots N\}$
$\quad t = 0$, $x^{(0)} = 0$, $\tau_0 = 0$
loop
$\quad t = t + 1$
$\quad \Lambda = \mathrm{supp}(x^{(t-1)}|_k)$, $\mathrm{w}_i = \begin{cases} \omega, & i \in \Lambda \\ 1, & i \in \Lambda^c \end{cases}$
$\quad \tau_{t-1}' = \|x^{(t-1)}\|_{1,\mathrm{w}}$
$\quad \tau_t = \tau_{t-1}' + \frac{\|r^{\tau_{t-1}}\|_2 \quad - \quad \epsilon}{\|A^H r^{\tau_{t-1}}\|_{\infty,\mathrm{w}} / \|r^{\tau_{t-1}}\|_2}$
$\quad x^{(t)} = \arg\min_u \|Au - y\|_2$ s.t. $\|u\|_{1,\mathrm{w}} \leq \tau_t$
$\quad r^{\tau_t} = y - Ax^{(t)}$
end loop

3.2. Discussion

The WSPGL1 algorithm converges to the solution of a weighted BPDN problem with weights $\omega \in [0,1]$ applied to a support set Λ. When the sparse signal is recovered exactly, the set Λ coincides with the true support of the sparse signal x. Figure 2 (a) illustrates the solution path of WSPGL1 which follows the Pareto curve of the BPDN problem until the first (LS_τ) is solved. The algorithm then uses the support information from x^{τ_1} to switch to the Pareto curve of



(a)



(b)

Fig. 2: (a) The solution path for WSPGL1 follows the BPDN Pareto curve until the first (LS_τ) is solved, after which WSPGL1 switches to the wBPDN Pareto curve. (b) Solution paths of WSPGL1, SPGL1, and weighted SPGL1 with oracle support information. Both WSPGL1 and the oracle weighted SPGL1 use $\omega = 0.3$.

the wBPDN problem. Figure 2 (b) compares the solution paths of WSPGL1, SPGL1, and oracle weighted SPGL1 with weight $\omega = 0.3$ applied to the true signal support. It can be seen that WSPGL1 converges to the solution of the oracle weighted ℓ_1 problem. Moreover, the solution paths of these algorithms merge after only the first (LS_τ) subproblem. Note here that the x-axis is the parameter τ which is equal to the one-norm of $x^{(t)}$ for SPGL1 and the weighted one-norm of $x^{(t)}$ for WSPGL1 and the oracle weighted SPGL1.

It is still not clear under what conditions the WSPGL1 algorithm achieves exact recovery. What is clear is that WSPGL1 can exactly recover signals with far more nonzero coefficients than what BPDN can recover. The WSPGL1 algorithm is motivated by the work in [9] and [12], which show that weighted ℓ_1 minimization can recover less sparse signals than BPDN when the weights are applied to a support

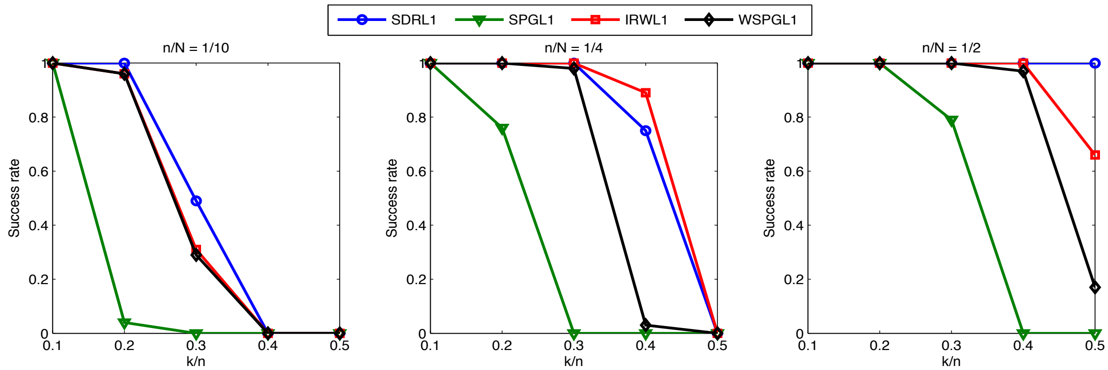


Fig. 3: Comparison of the percentage of exact recovery of *sparse signals* between the proposed WSPGL1, SDRL1 [12], IRL1 [10], and standard ℓ_1 minimization using SPGL1 [13]. The signals have an ambient dimension $N = 2000$ and the sparsity and number of measurements are varied. The results are averaged over 100 experiments.

estimate that is at least 50% accurate. Moreover, it is possible to draw a support estimate from the solution of BPDN and improve that support estimate by solving wBPDN using the initial support estimate. Based on these results, we conjectured that the solution of every LASSO subproblem in SPGL1 allows us to find a support estimate that is accurate enough to improve the recovery conditions of the corresponding wBPDN problem. A full analysis of this algorithm will be the subject of future work.

4. NUMERICAL RESULTS

We tested the WSPGL1 algorithm by comparing its performance with SDRL1 [12], IRWL1 [10] and standard ℓ_1 minimization using the SPGL1 [13] algorithm in recovering synthetic signals x of dimension $N = 2000$. We first recover sparse signals from compressed measurements $y = Ax$ using matrices A with i.i.d. Gaussian random entries and dimensions $n \times N$ where $n \in \{N/10, N/4, N/2\}$. The sparsity of the signal is varied such that $k/n \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. To quantify the reconstruction performance, we plot in Figure 3 the percentage of successful recovery averaged over 100 realizations of the same experimental conditions. The figure shows that in all cases, the WSPGL1 algorithm outperforms standard ℓ_1 minimization in recovering sparse signals. Moreover, the recovery performance approaches that of the iterative reweighted ℓ_1 algorithms SDRL1 and IRWL1 while requiring only a fraction of the computational cost associated with these algorithms.

5. REFERENCES

[1] D. Donoho, "Compressed sensing.," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, 2006.

[3] S. Chen, D. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.

[4] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335–355, May 2007.

[5] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 035020, 2008.

[6] R. Saab and O. Yilmaz, "Sparse recovery by non-convex optimization – instance optimality," *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 30–48, July 2010.

[7] R. von Borries, C.J. Miosso, and C. Potes, "Compressed sensing using prior information," in *2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMPSAP 2007.*, 12-14 2007, pp. 121 – 124.

[8] N. Vaswani and Wei Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *arXiv:0903.5066v4*, 2009.

[9] M. P. Friedlander, H. Mansour, R. Saab, and Ö. Yılmaz, "Recovering compressively sampled signals using partial support information," *to appear in the IEEE Trans. on Inf. Theory.*

[10] E. J. Candès, Michael B. Wakin, and Stephen P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[11] D. Needell, "Noisy signal recovery via iterative reweighted l1-minimization," in *Proceedings of the 43rd Asilomar conference on Signals, systems and computers*, 2009, Asilomar'09, pp. 113–117.

[12] H. Mansour and O. Yilmaz, "Support driven reweighted ℓ_1 minimization," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2012.

[13] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[14] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.